

dr hab. inż. Piotr Napieralski, prof. uczelni
Instytut Informatyki I72
Wydział Fizyki Technicznej Informatyki
i Matematyki Stosowanej
Politechnika Łódzka

Łódź 24.12.2022

Recenzja Rozprawy Doktorskiej

TYTUŁ ROZPRAWY:

Detekcja obiektów w obrazach cyfrowych z użyciem uczenia głębokiego w warunkach stratnej kompresji obrazu

AUTOR ROZPRAWY:

mgr inż. Tomasz Gandor

PROMOTOR ROZPRAWY:

dr hab. inż. Henryk Josiński

PROMOTOR POMOCNICZY ROZPRAWY:

dr hab. inż. Jakub Nalepa

Formalną podstawą opracowania recenzji jest pismo z dnia 06.10.2022, Przewodniczącej Rady Naukowej Dyscypliny Informatyki prof. dr hab. Marii Elżbiety Orłowskiej.

ZAKRES TEMATYCZNY I REALIZACJA ROZPRAWY

Rozprawa składa się z 8 rozdziałów. Całość rozprawy została przedstawiona na 155 stronach. W pracy Autor rozważa ważne zagadnienie dotyczące skuteczności detekcji obiektów w obrazach cyfrowych z użyciem metod uczenia głębokiego w warunkach stratnej kompresji obrazu. Tematyka pracy dotyczy istotnego problemu pojawiającego się bardzo często w praktyce zastosowania metod detekcji i klasyfikacji obiektów w obrazach cyfrowych z użyciem technik uczenia głębokiego, tzn. spadku skuteczności wspomnianych metod podczas realizacji procesu rozpoznawania obrazów poddanych kompresji. Autor trafnie skupił swoją uwagę na zbadaniu wpływu kompresji obrazów algorytmem JPEG na skuteczność detekcji i klasyfikacji obiektów, gdyż taki scenariusz jest stosunkowo najczęściej spotykanym w zastosowaniach praktycznych.

Rozdział pierwszy pracy zawiera wprowadzenie, które w ogólny sposób definiuje zakres pojęciowy poszczególnych zagadnień związanych z badaniami wykonanymi przez Autora w ramach pracy, tzn. (i) pojęcia detekcji obiektów na obrazach cyfrowych, (ii) pojęcia uczenia głębokiego, (iii) miar skuteczności detekcji obiektów oraz (iv) kompresji obrazów cyfrowych. W rozdziale Autor sprecyzował też cel pracy oraz sformułował trzy tezy pracy i wyjaśnił wkład swojej pracy we wzrost wiedzy dotyczącej tematyki skuteczności detekcji obiektów w obrazach cyfrowych z użyciem metod uczenia głębokiego w warunkach stratnej kompresji obrazu.

Rozdział drugi zawiera obszerny przegląd literatury dotyczącej rozważanego w pracy zagadnienia.

Rozdziały trzeci, czwarty i piąty mają charakter teoretyczny. Autor objaśnia w nich w sposób pogłębiony tematykę dotyczącą (i) uczenia maszynowego głębokiego, w szczególności z wykorzystaniem techniki transferu uczenia, (ii) detekcji obiektów na obrazach cyfrowych z użyciem algorytmów uczenia maszynowego głębokiego, w tym architektur modeli głębokich oraz miar skuteczności detekcji obiektów na obrazach cyfrowych, a także (iii) kompresji obrazów algorytmem JPEG, ze szczególnym uwzględnieniem mechanizmu sterowania stopniem kompresji obrazu oraz miar jakości obrazów poddawanych kompresji. Rozdziały te stanowią istotną część rozprawy, ponieważ opisane w nich zagadnienia pozwalają odbiorcy opracowania w sposób nie budzący wątpliwości dokonać interpretacji scenariuszy, parametrów, warunków oraz wyników przeprowadzonych przez Autora i przedstawionych w dalszej części pracy eksperymentów badawczych.

Rozdział szósty stanowi główną część rozprawy doktorskiej, w której w sposób systematyczny Autor przedstawia szczegółowy opis badań eksperymentalnych, przeprowadzonych w trakcie realizacji rozprawy, wraz z ich rezultatami uzupełnionymi o wykresy i komentarze ułatwiające czytelnikowi interpretację przedstawionych wyników.

W rozdziale siódmym Autor dokonał omówienia wyników badań empirycznych zrealizowanych w ramach pracy, sformułował wnioski odnoszące się do postawionych tez, a także wskazał potencjalne możliwości poprawy skuteczności detekcji obiektów w obrazach cyfrowych z użyciem metod uczenia głębokiego w warunkach stratnej kompresji obrazu.

W Rozdziale ósmym znajduje się podsumowanie pracy, w którym Autor wskazuje zasadność tez pracy oraz przedstawia swój dorobek naukowy.

Rozprawa zawiera także spis rysunków oraz tabel, bibliografię oraz cztery dodatki zawierające wykaz skrótów i symboli użytych w opisach przedstawianych zagadnień, tabele miar skuteczności badanych modeli uczenia głębokiego oraz kody źródłowe funkcji detekcji obiektów i obliczania jej miar skuteczności. Ważną częścią pracy są także aktualizowane na bieżąco materiały źródłowe i wyniki badań empirycznych zebrane podczas realizacji rozprawy w postaci repozytoriów publicznych.

OMÓWIENIE TREŚCI I WYNIKÓW ROZPRAWY

Zgodnie z celami pracy oraz Autor przeprowadził badania eksperymentalne skuteczności detekcji obiektów w obrazach cyfrowych z użyciem metod uczenia głębokiego w warunkach stratnej kompresji obrazu. Przedstawione w rozdziale siódmym rozprawy wnioski z badań eksperymentalnych dowiodły, zdaniem recenzenta, słuszności jej tez.

Główną częścią pracy i zarazem głównym wkładem Autora w poszerzenie wiedzy na temat zagadnienia skuteczności detekcji obiektów w obrazach cyfrowych z użyciem metod uczenia głębokiego w warunkach stratnej kompresji obrazu jest jej część badawcza opisana w rozdziale szóstym. Autor podzielił eksperymenty dotyczące

modeli uczenia głębokiego na dwa główne etapy. W obydwu etapach do analizy modeli detekcji obiektów posłużono się dostępnym publicznie zbiorem obrazów i metadanych COCO, składającym się z 5000 nie-ikonicznych obrazów zawierających 80 klas obiektów do wykrycia.

W etapie pierwszym Autor poddał badaniom skuteczność detekcji obiektów w funkcji współczynnika kompresji obrazów dla wybranych dostępnych publicznie dziewięciu pretrenowanych modeli uczenia głębokiego. W treningu modeli etapu pierwszego nie brały udziału obrazy skompresowane, pomijając fakt niewielkiej zastanej degradacji jakości obrazów źródłowych wynikający z ich oryginalnego zapisu w formacie JFIF (pliki z rozszerzeniami jpg lub jpeg).

W etapie drugim Autor dokonał próby identyfikacji możliwości poprawy skuteczności detekcji obiektów w obrazach cyfrowych w warunkach kompresji stratnej poprzez dodatkowe uwzględnienie w wybranych scenariuszach treningowych obrazów skompresowanych z dwoma współczynnikami jakości kompresji Q algorytmu JPEG. W etapie tym Autor poddał analizie osiem modeli detekcji o dwóch różnych typach architektur, w tym modeli pre-trenowanych jak i tych, których trening rozpoczynano od losowych wartości wagowych. Wybór takiej konfiguracji eksperymentalnej został uzasadniony przez Autora faktem, iż odpowiada ona rzeczywistym scenariuszom treningowym uwzględniającym dostępne zasoby sprzętowe.

Dla obydwu etapów badawczych Autor dokonał opisu warunków eksperymentów umożliwiającego ich replikację oraz podał wyniki dotyczące wartości miar skuteczności detekcji - m.in. precyzji, czułości, miary F1 oraz rodziny miar średniej precyzji AP, w funkcji współczynnika jakości kompresji Q algorytmu JPEG.

Wyniki uzyskane w obydwu etapach eksperymentalnych zostały następnie omówione w rozdziale siódmym, w którym zaproponowano także, bazując na poczynionych obserwacjach i inspirowane dodatkowo źródłami bibliograficznymi, usprawnienia procedur treningowych rozważanych w pracy detektorów potencjalnie zwiększające szeroko rozumianą skuteczność ich zastosowania w praktyce.

Autor dokonuje ostatecznej syntezy i podsumowania wyników pracy w rozdziale ósmym, stwierdzając słuszność postawionych przez siebie tez.

Przedstawiona praca ma, zdaniem recenzenta, wysoką wartość poznawczą dotyczącą rozważanego zagadnienia, szczególnie w aspekcie praktycznym. Przedstawione w rozprawie wyniki eksperymentów, uzupełnione o udostępnione przez Autora w formie źródłowej materiały zawarte w publicznie dostępnych repozytoriach internetowych, stanowią cenną bazę wiedzy dla praktyków detekcji obiektów w obrazach cyfrowych w warunkach kompresji stratnej.

Postawione przez Autora na str. 7 i 8 rozprawy trzy tezy pracy zostały zdaniem recenzenta udowodnione w kontekście zaprezentowanych w rozdziałach 6-8 oraz w dodatkach A-C wyników i ich dyskusji przeprowadzonej przez Autora w rozdziałach 6-8.

UWAGI POZYTYWNE

Jak wspomniano w poprzednim podpunkcie główną wartością przedstawionej do recenzji pracy jest, zdaniem recenzenta, bogaty zestaw wyników badań eksperymentalnych, który może służyć jako cenna, gotowa do praktycznego wykorzystania, baza wiedzy dla specjalistów z dziedziny detekcji obiektów w obrazach cyfrowych w warunkach kompresji stratnej. Treść pracy została przedstawiona w taki sposób, że umożliwia ekspertowi z dziedziny detekcji obiektów sprawną identyfikację architektury i specyficznej konfiguracji detektora do zadania rozpoznawania przy wybranych warunkach sprzętowych i tych dotyczących stopnia degradacji obrazów poddawanych procesowi detekcji spowodowanych ich kompresją. Jest to niewątpliwie wysoce pozytywna cecha przedstawionej rozprawy.

Recenzent chce także zwrócić uwagę na stronę językową rozprawy, którą ocenia bardzo wysoko ze względu na klarowność, a jednocześnie zwięzłość opisu poszczególnych zagadnień obecnych w rozprawie. Podobnie wysoko należy ocenić jakość strony typograficznej rozprawy.

Sformułowana wyżej wysoce pozytywna ocena przedstawionej do recenzji rozprawy, nie oznacza jednak, że zdaniem recenzenta, Autor nie ustrzegł się pewnych niedogodności dotyczących zagadnień merytorycznych, kwestii organizacji prezentacji treści pracy, a także drobnych błędów językowych czy typograficznych. Zostaną one wyszczególnione w formie uwag, sugestii i pytań do Autora w kolejnym podpunkcie niniejszej recenzji.

UWAGI DYSKUSYJNE I KRYTYCZNE

W niniejszym podpunkcie sformułowane zostaną uwagi dyskusyjne, krytyczne, sugestie oraz pytania recenzenta do Autora przedstawionej rozprawy. Są one wyszczególnione niżej.

UWAGI MERYTORYCZNE

1. W podrozdziale 5.2.1 na str. 65 rozprawy Autor podał przykładowe tablice kwantyzacji algorytmu JPEG dla składowych luminancji i chrominancji obrazu dla wybranych wartości parametru jakości Q. Zgodnie z opisem na str. 75 do kompresji obrazów podczas testów detekcji Autor używa polecenia mogrify z pakietu ImageMagick. Czy Autor ma pewność, że polecenie mogrify korzysta z tych tablic? Gdyby tak nie było, mogłoby to mieć pewien wpływ na zakresy wartości parametru Q, które raportowane są w rozprawie przy podawaniu wyników testów detekcji obiektów.
2. W tabeli 6.2 w podrozdziale 6.2 na str. 79 rozprawy Autor podaje statystyki dotyczące wartości współczynnika kompresji obrazów w % w zależności od wartości parametru jakości kompresji Q algorytmu JPEG, przy czym definicja współczynnika kompresji, podana przez Autora na stronie 78, określona jest jako odwrotność stopnia kompresji będącego ilorazem rozmiaru danych nieskompresowanych do skompresowanych. Recenzent ma wątpliwości co wartości podanych w tabeli 6.2, np. jeśli stopień kompresji obrazu wynosiłby 2, wówczas współczynnik kompresji miałby wartość 0.5, a zatem rozmiar obrazu skompresowanego stanowiłby 50%

rozmiaru obrazu oryginalnego, wobec tego wartość współczynnika kompresji wyrażona w %, wg definicji ze strony 78, wynosiłaby 50. A zatem podobna wartość powinna znajdować się pośród wartości zawartych w ta-beli 6.2, tymczasem podobne wartości nie są tam obecne. Recenzent prosi Autora o wy-jaśnienie tej kwestii.

3. Pewną trudność interpretacyjną sprawia czytelnikowi organizacja prezentacji wyników eksperymentalnych w etapie I badań empirycznych. Z jednej strony wyniki bazowe miar zależnych od wyboru progu ufności oraz miar średniej precyzji podane są w postaci szczegółowej, z użyciem tabel. Z drugiej zaś, najistotniejsze z punktu widzenia odbiorcy pracy, wyniki wspomnianych miar dla poszczególnych wartości parametru jakości obrazów skompresowanych Q są podane w postaci szeregu wykresów, jednak bez bezpośredniego odniesienia się Autora do kwestii ich interpretacji w kontekście zaprezentowanych wcześniej wyników bazowych. Sprawia to wrażenie jakoby bazowe wyniki miar, jako takie, miały szczególnie istotne znaczenie dla całości wyводу zawartego w rozprawie w porównaniu do wyników sparametryzowanych współczynnikiem jakości Q . Mając świadomość możliwości błędnej oceny przez Recenzenta wspomnianej tu kwestii, prosi on o wyjaśnienie motywacji Autora stojącej za takim wyborem organizacji prezentacji wyników etapu I.
4. W etapie II badań empirycznych Autor wybrał do testowania osiem modeli o dwóch różnych architekturach, trenowanych w pełni i pre-trenowanych na zbiorze zawierającym obrazy skompresowane przy założeniu dwóch scenariuszy dotyczących wydajności zasobów sprzętowych. Jak wspomina Autor, zawężenie wyboru trenowanych modeli wynikało ze znacznych wymagań obliczeniowych etapu II, a co za tym idzie czasu wymaganego do przeprowadzenia ich treningu. Jednak Autor nie wyjaśnił motywacji stojącej za wyborem tych konkretnych modeli spośród wszystkich, które rozważane były w etapie I. Stąd pytanie recenzenta - jaka była motywacja Autora stojąca za wyborem modeli analizowanych w etapie II badań empirycznych?
5. Zdaniem recenzenta istotną niedogodnością w interpretacji wyników etapu II badań empirycznych jest fakt braku ich odniesienia ilościowego do analogicznych wyników uzyskanych dla odpowiadających im modeli detektorów z etapu I. Powoduje to trudność w porównaniu wyników skuteczności detekcji obiektów uzyskanych dla odpowiadających sobie modeli, które były trenowane na zbiorach nie zawierających obrazów skompresowanych z tymi, które takie obrazy zawierały, a co za tym idzie, trudność weryfikacji trzeciej tezy rozprawy. Recenzent prosi Autora o wyjaśnienie tej kwestii, być może zasugerowane przez recenzenta porównanie nie byłoby z jakich względów uprawnione.

UWAGI TYPOGRAFICZNE I JĘZYKOWE

1. Wykresy miar skuteczności detekcji obiektów zależnych od wyboru progu ufności oraz miar średniej precyzji mogłyby być zdaniem recenzenta nieco większe. Pozwoliłoby to np. na rezygnację z tabelarycznej formy prezentacji wartości bazowych miar skuteczności detekcji na rzecz możliwości przybliżonego ich

odczytu ze wspomnianych wykresów, co w świetle uwagi nr 3 zawartej w poprzednim paragrafie, mogłoby usunąć wrażenie, iż wartości bazowe miar skuteczności detekcji mają szczególnie istotne znaczenie dla całości wyводу zawartego w rozprawie w porównaniu do wyników miar sparametryzowanych współczynnikiem jakości Q.

2. Zwiększenie rozmiaru wykresów skutkowałoby również dalszą poprawą czytelności niektórych z nich, np. wykresu 6.10, gdzie dobór kolorów powoduje trudność w rozróżnieniu miar, których dotyczą poszczególne linie wykresu, czy np. wykresu 6.11, w którym z kolei różnice grubości linii są na tyle nieznaczące, że istotnie utrudniają podobne rozróżnienie.
3. Podczas lektury rozprawy recenzent nie odnotował błędów językowych, poza jednym przypadkiem: na str. 107 w trzeciej linii pierwszego paragrafu za listą punktowaną znajduje się zdanie: „Miara AP jest średnią arytmetyczną dziesięciu miar, od AP50 do AP95, w miara AP75 prawie na środkową z nich.”, które ewidentnie jest wynikiem pomyłki językowej Autora.

PODSUMOWANIE

Podsumowując, rozprawa stanowi cenny wkład Autora w poszerzenie wiedzy dotyczącej dziedziny detekcji obiektów w obrazach cyfrowych w warunkach kompresji stratnej, co jest wynikiem oryginalnym. Wymienione przez Autora w podsumowaniu rozprawy fakty publikacji części jej wyników ich prezentacja na konferencjach naukowych oznacza, że wyniki te zostały dodatkowo pozytywnie zweryfikowane przez środowisko naukowe. Uwzględniając zarówno uwagi pozytywne jak i te dyskusyjne i krytyczne oraz biorąc pod uwagę treść niniejszej recenzji, stwierdzam, że recenzowana rozprawa stanowi oryginalne rozwiązanie problemu naukowego. Wykazuje ona także ogólną wiedzę teoretyczną kandydata w przedmiotowej dyscyplinie naukowej oraz umiejętność samodzielnego prowadzenia pracy naukowej.

Stwierdzam, że opiniowana rozprawa doktorska mgr inż. Tomasza Gandora pod tytułem „Detekcja obiektów w obrazach cyfrowych z użyciem uczenia głębokiego w warunkach stratnej kompresji obrazu” spełnia wymogi formalne. Wnoszę o dopuszczenie przedstawionej do recenzji rozprawy doktorskiej do dalszych etapów przewodu doktorskiego.

Piotr Napiewałski

